In [1]:
```
import pandas as pd
import numpy as np

df=pd.read_csv("/home/student/Desktop/COTB21/Academic_performance.c
```

In [2]:
```
df
```

Out[2]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	86.0	73.0	100.0	2018.0	1.0	Other
1	78.0	90.0	66.0	79.0	2018.0	2.0	Male
2	77.0	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.0	NaN	74.0	88.0	2019.0	3.0	Male
4	74.0	92.0	90.0	68.0	2018.0	3.0	Other
5	72.0	94.0	72.0	98.0	2020.0	NaN	Female
6	NaN	80.0	71.0	88.0	NaN	3.0	Other
7	40.0	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.0	78.0	69.0	92.0	2020.0	3.0	Other
9	78.0	82.0	63.0	86.0	2019.0	3.0	Other
10	77.0	87.0	66.0	79.0	2019.0	2.0	Male
11	79.0	NaN	64.0	NaN	2018.0	1.0	Other
12	66.0	82.0	79.0	93.0	2019.0	3.0	Male
13	74.0	84.0	69.0	86.0	2020.0	3.0	Male
14	72.0	89.0	75.0	83.0	2020.0	NaN	Male
15	NaN	78.0	NaN	97.0	2019.0	3.0	Other
16	NaN	76.0	74.0	98.0	NaN	3.0	Female
17	67.0	91.0	75.0	87.0	2018.0	3.0	Male
18	65.0	75.0	NaN	102.0	2019.0	3.0	Other
19	67.0	NaN	64.0	83.0	2021.0	2.0	NaN
20	71.0	NaN	60.0	82.0	2023.0	2.0	Female
21	74.0	87.0	78.0	NaN	2020.0	1.0	Male
22	89.0	89.0	66.0	97.0	2019.0	3.0	Female
23	70.0	84.0	72.0	84.0	2019.0	2.0	Male
24	61.0	89.0	69.0	99.0	2020.0	3.0	Other
25	60.0	78.0	10.0	82.0	2020.0	2.0	Female
26	72.0	50.0	80.0	99.0	2021.0	3.0	Female
27	62.0	76.0	73.0	98.0	2020.0	3.0	Male
28	74.0	88.0	74.0	94.0	2019.0	3.0	Other
29	66.0	86.0	62.0	94.0	2018.0	3.0	Other

In [3]:
```
df.isnull()
```

Out[3]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	True	True	False	False	True
3	False	True	False	False	False	False	False
4	False	False	False	False	False	False	False
5	False	False	False	False	False	True	False
6	True	False	False	False	True	False	False
7	False	False	False	False	False	False	True
8	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False
11	False	True	False	True	False	False	False
12	False	False	False	False	False	False	False
13	False	False	False	False	False	False	False
14	False	False	False	False	False	True	False
15	True	False	True	False	False	False	False
16	True	False	False	False	True	False	False
17	False	False	False	False	False	False	False
18	False	False	True	False	False	False	False
19	False	True	False	False	False	False	True
20	False	True	False	False	False	False	False
21	False	False	False	True	False	False	False
22	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False
25	False	False	False	False	False	False	False
26	False	False	False	False	False	False	False
27	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False
29	False	False	False	False	False	False	False

In [4]:
```
series= pd.isnull(df("math score"))
df[series]
```

```
---------------------------------------------------------------------
----------
TypeError                                 Traceback (most recent
call last)
Cell In[4], line 1
----> 1 series= pd.isnull(df("math score"))
      2 df[series]

TypeError: 'DataFrame' object is not callable
```

In [5]:
```
series= pd.isnull(df["math score"])
df[series]
```

Out[5]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
6	NaN	80.0	71.0	88.0	NaN	3.0	Other
15	NaN	78.0	NaN	97.0	2019.0	3.0	Other
16	NaN	76.0	74.0	98.0	NaN	3.0	Female

In [6]:
```
df.notnull()
```

Out[6]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True
2	True	True	False	False	True	True	False
3	True	False	True	True	True	True	True
4	True	True	True	True	True	True	True
5	True	True	True	True	True	False	True
6	False	True	True	True	False	True	True
7	True	True	True	True	True	True	False
8	True	True	True	True	True	True	True
9	True	True	True	True	True	True	True
10	True	True	True	True	True	True	True
11	True	False	True	False	True	True	True
12	True	True	True	True	True	True	True
13	True	True	True	True	True	True	True
14	True	True	True	True	True	False	True
15	False	True	False	True	True	True	True
16	False	True	True	True	False	True	True
17	True	True	True	True	True	True	True
18	True	True	False	True	True	True	True
19	True	False	True	True	True	True	False
20	True	False	True	True	True	True	True

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
21	True	True	True	False	True	True	True
22	True	True	True	True	True	True	True
23	True	True	True	True	True	True	True
24	True	True	True	True	True	True	True
25	True	True	True	True	True	True	True
26	True	True	True	True	True	True	True
27	True	True	True	True	True	True	True

In [7]:
```
series1 = pd.notnull(df["math score"])
df[series1]
```

Out[7]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	86.0	73.0	100.0	2018.0	1.0	Other
1	78.0	90.0	66.0	79.0	2018.0	2.0	Male
2	77.0	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.0	NaN	74.0	88.0	2019.0	3.0	Male
4	74.0	92.0	90.0	68.0	2018.0	3.0	Other
5	72.0	94.0	72.0	98.0	2020.0	NaN	Female
7	40.0	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.0	78.0	69.0	92.0	2020.0	3.0	Other
9	78.0	82.0	63.0	86.0	2019.0	3.0	Other
10	77.0	87.0	66.0	79.0	2019.0	2.0	Male
11	79.0	NaN	64.0	NaN	2018.0	1.0	Other
12	66.0	82.0	79.0	93.0	2019.0	3.0	Male
13	74.0	84.0	69.0	86.0	2020.0	3.0	Male
14	72.0	89.0	75.0	83.0	2020.0	NaN	Male
17	67.0	91.0	75.0	87.0	2018.0	3.0	Male
18	65.0	75.0	NaN	102.0	2019.0	3.0	Other
19	67.0	NaN	64.0	83.0	2021.0	2.0	NaN
20	71.0	NaN	60.0	82.0	2023.0	2.0	Female
21	74.0	87.0	78.0	NaN	2020.0	1.0	Male
22	89.0	89.0	66.0	97.0	2019.0	3.0	Female
23	70.0	84.0	72.0	84.0	2019.0	2.0	Male
24	61.0	89.0	69.0	99.0	2020.0	3.0	Other
25	60.0	78.0	10.0	82.0	2020.0	2.0	Female
26	72.0	50.0	80.0	99.0	2021.0	3.0	Female

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
27	62.0	76.0	73.0	98.0	2020.0	3.0	Male

In [8]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['gender'] = le.fit_transform(df['gender'])
newdf=df
df
```

Out[8]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	86.0	73.0	100.0	2018.0	1.0	2
1	78.0	90.0	66.0	79.0	2018.0	2.0	1
2	77.0	69.0	NaN	NaN	2018.0	1.0	3
3	73.0	NaN	74.0	88.0	2019.0	3.0	1
4	74.0	92.0	90.0	68.0	2018.0	3.0	2
5	72.0	94.0	72.0	98.0	2020.0	NaN	0
6	NaN	80.0	71.0	88.0	NaN	3.0	2
7	40.0	82.0	63.0	77.0	2015.0	2.0	3
8	62.0	78.0	69.0	92.0	2020.0	3.0	2
9	78.0	82.0	63.0	86.0	2019.0	3.0	2
10	77.0	87.0	66.0	79.0	2019.0	2.0	1
11	79.0	NaN	64.0	NaN	2018.0	1.0	2
12	66.0	82.0	79.0	93.0	2019.0	3.0	1
13	74.0	84.0	69.0	86.0	2020.0	3.0	1
14	72.0	89.0	75.0	83.0	2020.0	NaN	1
15	NaN	78.0	NaN	97.0	2019.0	3.0	2
16	NaN	76.0	74.0	98.0	NaN	3.0	0
17	67.0	91.0	75.0	87.0	2018.0	3.0	1
18	65.0	75.0	NaN	102.0	2019.0	3.0	2
19	67.0	NaN	64.0	83.0	2021.0	2.0	3
20	71.0	NaN	60.0	82.0	2023.0	2.0	0
21	74.0	87.0	78.0	NaN	2020.0	1.0	1
22	89.0	89.0	66.0	97.0	2019.0	3.0	0
23	70.0	84.0	72.0	84.0	2019.0	2.0	1
24	61.0	89.0	69.0	99.0	2020.0	3.0	2
25	60.0	78.0	10.0	82.0	2020.0	2.0	0
26	72.0	50.0	80.0	99.0	2021.0	3.0	0
27	62.0	76.0	73.0	98.0	2020.0	3.0	1

	math score	reading score	writing score	placement score	club join year	placement offer count	gender

In [ ]:

In [10]:
```
missing_values=["Na","na"]
df=pd.read_csv("/home/student/Desktop/COTB21/Academic_performance.c
df
```

Out[10]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	86.0	73.0	100.0	2018.0	1.0	Other
1	78.0	90.0	66.0	79.0	2018.0	2.0	Male
2	77.0	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.0	NaN	74.0	88.0	2019.0	3.0	Male
4	74.0	92.0	90.0	68.0	2018.0	3.0	Other
5	72.0	94.0	72.0	98.0	2020.0	NaN	Female
6	NaN	80.0	71.0	88.0	NaN	3.0	Other
7	40.0	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.0	78.0	69.0	92.0	2020.0	3.0	Other
9	78.0	82.0	63.0	86.0	2019.0	3.0	Other
10	77.0	87.0	66.0	79.0	2019.0	2.0	Male
11	79.0	NaN	64.0	NaN	2018.0	1.0	Other
12	66.0	82.0	79.0	93.0	2019.0	3.0	Male
13	74.0	84.0	69.0	86.0	2020.0	3.0	Male
14	72.0	89.0	75.0	83.0	2020.0	NaN	Male
15	NaN	78.0	NaN	97.0	2019.0	3.0	Other
16	NaN	76.0	74.0	98.0	NaN	3.0	Female
17	67.0	91.0	75.0	87.0	2018.0	3.0	Male
18	65.0	75.0	NaN	102.0	2019.0	3.0	Other
19	67.0	NaN	64.0	83.0	2021.0	2.0	NaN
20	71.0	NaN	60.0	82.0	2023.0	2.0	Female
21	74.0	87.0	78.0	NaN	2020.0	1.0	Male
22	89.0	89.0	66.0	97.0	2019.0	3.0	Female
23	70.0	84.0	72.0	84.0	2019.0	2.0	Male
24	61.0	89.0	69.0	99.0	2020.0	3.0	Other
25	60.0	78.0	10.0	82.0	2020.0	2.0	Female
26	72.0	50.0	80.0	99.0	2021.0	3.0	Female
27	62.0	76.0	73.0	98.0	2020.0	3.0	Male
28	74.0	88.0	74.0	94.0	2019.0	3.0	Other

	math	reading	writing	placement	club join	placement offer	.

In [12]:
```
ndf=df
ndf.fillna(0)
```

Out[12]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	86.0	73.0	100.0	2018.0	1.0	Other
1	78.0	90.0	66.0	79.0	2018.0	2.0	Male
2	77.0	69.0	0.0	0.0	2018.0	1.0	0
3	73.0	0.0	74.0	88.0	2019.0	3.0	Male
4	74.0	92.0	90.0	68.0	2018.0	3.0	Other
5	72.0	94.0	72.0	98.0	2020.0	0.0	Female
6	0.0	80.0	71.0	88.0	0.0	3.0	Other
7	40.0	82.0	63.0	77.0	2015.0	2.0	0
8	62.0	78.0	69.0	92.0	2020.0	3.0	Other
9	78.0	82.0	63.0	86.0	2019.0	3.0	Other
10	77.0	87.0	66.0	79.0	2019.0	2.0	Male
11	79.0	0.0	64.0	0.0	2018.0	1.0	Other
12	66.0	82.0	79.0	93.0	2019.0	3.0	Male
13	74.0	84.0	69.0	86.0	2020.0	3.0	Male
14	72.0	89.0	75.0	83.0	2020.0	0.0	Male
15	0.0	78.0	0.0	97.0	2019.0	3.0	Other
16	0.0	76.0	74.0	98.0	0.0	3.0	Female
17	67.0	91.0	75.0	87.0	2018.0	3.0	Male
18	65.0	75.0	0.0	102.0	2019.0	3.0	Other
19	67.0	0.0	64.0	83.0	2021.0	2.0	0
20	71.0	0.0	60.0	82.0	2023.0	2.0	Female
21	74.0	87.0	78.0	0.0	2020.0	1.0	Male
22	89.0	89.0	66.0	97.0	2019.0	3.0	Female
23	70.0	84.0	72.0	84.0	2019.0	2.0	Male
24	61.0	89.0	69.0	99.0	2020.0	3.0	Other
25	60.0	78.0	10.0	82.0	2020.0	2.0	Female
26	72.0	50.0	80.0	99.0	2021.0	3.0	Female
27	62.0	76.0	73.0	98.0	2020.0	3.0	Male
28	74.0	88.0	74.0	94.0	2019.0	3.0	Other
29	66.0	86.0	62.0	94.0	2018.0	3.0	Other

In [13]:
```
m_v=df['math score'].mean()
df['math score'].fillna(value=m_v, inplace=True)
```

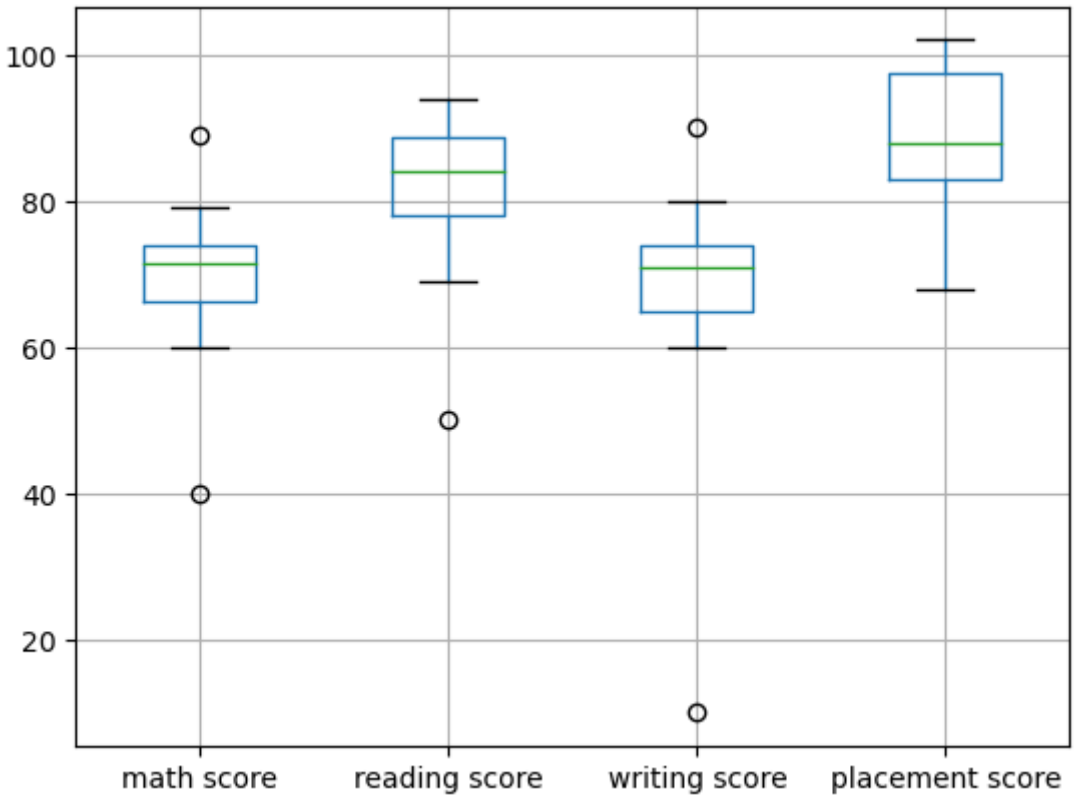
Out[13]:

```
df
```

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.000000	86.0	73.0	100.0	2018.0	1.0	Other
1	78.000000	90.0	66.0	79.0	2018.0	2.0	Male
2	77.000000	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.000000	NaN	74.0	88.0	2019.0	3.0	Male
4	74.000000	92.0	90.0	68.0	2018.0	3.0	Other
5	72.000000	94.0	72.0	98.0	2020.0	NaN	Female
6	70.074074	80.0	71.0	88.0	NaN	3.0	Other
7	40.000000	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.000000	78.0	69.0	92.0	2020.0	3.0	Other
9	78.000000	82.0	63.0	86.0	2019.0	3.0	Other
10	77.000000	87.0	66.0	79.0	2019.0	2.0	Male
11	79.000000	NaN	64.0	NaN	2018.0	1.0	Other
12	66.000000	82.0	79.0	93.0	2019.0	3.0	Male
13	74.000000	84.0	69.0	86.0	2020.0	3.0	Male
14	72.000000	89.0	75.0	83.0	2020.0	NaN	Male
15	70.074074	78.0	NaN	97.0	2019.0	3.0	Other
16	70.074074	76.0	74.0	98.0	NaN	3.0	Female
17	67.000000	91.0	75.0	87.0	2018.0	3.0	Male
18	65.000000	75.0	NaN	102.0	2019.0	3.0	Other
19	67.000000	NaN	64.0	83.0	2021.0	2.0	NaN
20	71.000000	NaN	60.0	82.0	2023.0	2.0	Female
21	74.000000	87.0	78.0	NaN	2020.0	1.0	Male
22	89.000000	89.0	66.0	97.0	2019.0	3.0	Female
23	70.000000	84.0	72.0	84.0	2019.0	2.0	Male
24	61.000000	89.0	69.0	99.0	2020.0	3.0	Other
25	60.000000	78.0	10.0	82.0	2020.0	2.0	Female
26	72.000000	50.0	80.0	99.0	2021.0	3.0	Female
27	62.000000	76.0	73.0	98.0	2020.0	3.0	Male
28	74.000000	88.0	74.0	94.0	2019.0	3.0	Other
29	66.000000	86.0	62.0	94.0	2018.0	3.0	Other

In [15]:
```
col = ['math score', 'reading score' , 'writing score','placement s
df.boxplot(col)
```

Out[15]: <Axes: >



In [16]:
```
print(np.where(df['math score']>90))
print(np.where(df['reading score']<25))
print(np.where(df['writing score']<30))
```

```
(array([], dtype=int64),)
(array([], dtype=int64),)
(array([25]),)
```

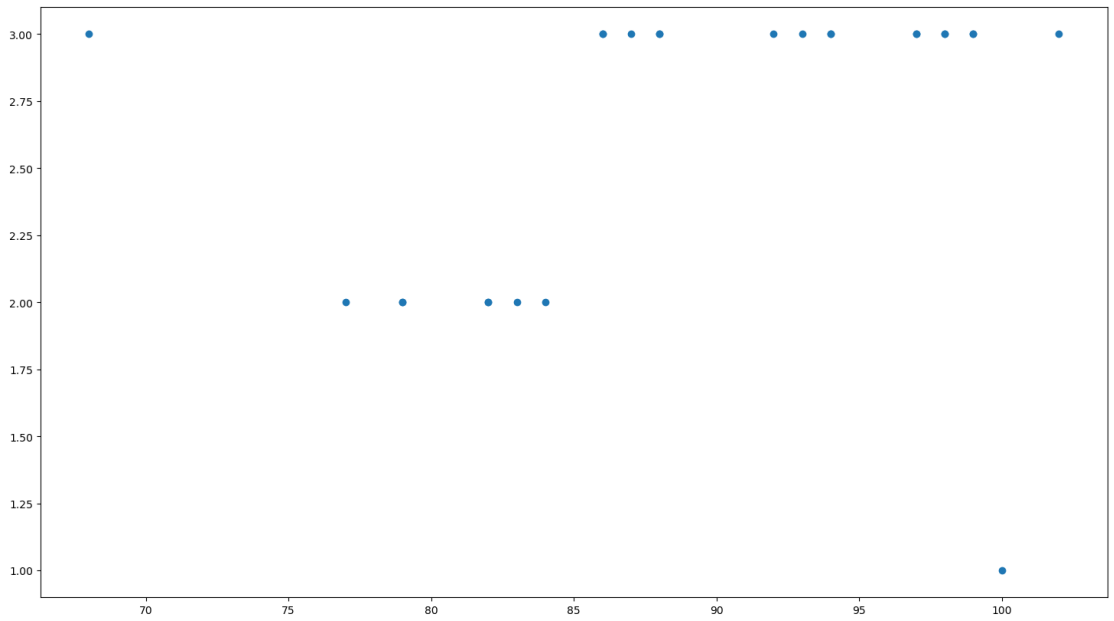
In [17]:
```
import matplotlib.pyplot as plt
df
```

Out[17]:

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.000000	86.0	73.0	100.0	2018.0	1.0	Other
1	78.000000	90.0	66.0	79.0	2018.0	2.0	Male
2	77.000000	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.000000	NaN	74.0	88.0	2019.0	3.0	Male
4	74.000000	92.0	90.0	68.0	2018.0	3.0	Other
5	72.000000	94.0	72.0	98.0	2020.0	NaN	Female
6	70.074074	80.0	71.0	88.0	NaN	3.0	Other
7	40.000000	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.000000	78.0	69.0	92.0	2020.0	3.0	Other
9	78.000000	82.0	63.0	86.0	2019.0	3.0	Other

	math score	reading score	writing score	placement score	club join year	placement offer count	gender
10	77.000000	87.0	66.0	79.0	2019.0	2.0	Male
11	79.000000	NaN	64.0	NaN	2018.0	1.0	Other
12	66.000000	82.0	79.0	93.0	2019.0	3.0	Male
13	74.000000	84.0	69.0	86.0	2020.0	3.0	Male
14	72.000000	89.0	75.0	83.0	2020.0	NaN	Male
15	70.074074	78.0	NaN	97.0	2019.0	3.0	Other
16	70.074074	76.0	74.0	98.0	NaN	3.0	Female
17	67.000000	91.0	75.0	87.0	2018.0	3.0	Male
18	65.000000	75.0	NaN	102.0	2019.0	3.0	Other
19	67.000000	NaN	64.0	83.0	2021.0	2.0	NaN
20	71.000000	NaN	60.0	82.0	2023.0	2.0	Female
21	74.000000	87.0	78.0	NaN	2020.0	1.0	Male
22	89.000000	89.0	66.0	97.0	2019.0	3.0	Female
23	70.000000	84.0	72.0	84.0	2019.0	2.0	Male
24	61.000000	89.0	69.0	99.0	2020.0	3.0	Other
25	60.000000	78.0	10.0	82.0	2020.0	2.0	Female
26	72.000000	50.0	80.0	99.0	2021.0	3.0	Female
27	62.000000	76.0	73.0	98.0	2020.0	3.0	Male

In [21]:
```
fig, ax =plt.subplots(figsize=(18,10))
ax.scatter(df['placement score'],df['placement offer count'])
plt.show()
```



In [22]:
```
print(np.where((df['placement score']<50) & (df['placement score']>
print(np.where((df['placement offer count']<3)))
```

```
(array([], dtype=int64),)
(array([ 0,  1,  2,  7, 10, 11, 19, 20, 21, 23, 25]),)
```

In [23]:
```
from scipy import stats
z=np.abs(stats.zscore(df['math score']))
print(z)
```

```
0     0.233447
1     0.960724
2     0.839511
3     0.354660
4     0.475873
5     0.233447
6     0.000000
7     3.645363
8     0.978681
9     0.960724
10    0.839511
11    1.081937
12    0.493830
13    0.475873
14    0.233447
15    0.000000
16    0.000000
17    0.372617
18    0.615043
19    0.372617
20    0.112234
21    0.475873
22    2.294065
23    0.008979
24    1.099894
25    1.221107
26    0.233447
27    0.978681
28    0.475873
29    0.493830
Name: math score, dtype: float64
```

In [25]:
```
threshold=0.18
sample_outliers=np.where(z<threshold)
sample_outliers
```

Out[25]:
```
(array([ 6, 15, 16, 20, 23]),)
```

In [26]:
```
sorted_rscore=sorted(df['reading score'])
sorted_rscore
```

Out[26]:

```
[50.0,
 69.0,
 75.0,
 76.0,
 76.0,
 78.0,
 78.0,
 78.0,
 86.0,
 90.0,
 nan,
 80.0,
 82.0,
 82.0,
 82.0,
 84.0,
 84.0,
 86.0,
 87.0,
 87.0
```

In [27]:
```
q1=np.percentile(sorted_rscore,25)
q3=np.percentile(sorted_rscore,75)
print(q1,q3)
```

```
nan nan
```

In [28]:
```
sorted_rscore=sorted(df['math score'])
sorted_rscore
```

Out[28]:
```
[40.0,
 60.0,
 61.0,
 62.0,
 62.0,
 65.0,
 66.0,
 66.0,
 67.0,
 67.0,
 70.0,
 70.07407407407408,
 70.07407407407408,
 70.07407407407408,
 71.0,
 72.0,
 72.0,
 72.0,
 72.0,
 73.0,
 74.0,
 74.0,
 74.0,
 74.0,
 77.0,
 77.0,
 78.0,
 78.0,
 79.0,
 89.0]
```

In [29]:
```
q1=np.percentile(sorted_rscore,25)
q3=np.percentile(sorted_rscore,75)
print(q1,q3)
```

```
66.25 74.0
```

In [32]:
```
IQR=q3-q1
lwr_bound=q1-(1.5*IQR)
upr_bound=q3+(1.5*IQR)
print(lwr_bound,upr_bound)
```

```
54.625 85.625
```

In [33]:
```
r_outliers=[]
for i in sorted_rscore:
    if(i<lwr_bound or i>upr_bound):
        r_outliers.append(i)
print(r_outliers)
```

```
[40.0, 89.0]
```

In [45]:
```
df_stud=df
ninetieth_percentile=np.percentile(df_stud['math score'], 90)
b = np.where(df_stud['math score']>ninetieth_percentile,ninetieth_p
print("New array:",b)
```

```
New array: [72. 78. 77. 73. 74. 72. 40. 62. 78. 77. 78. 66. 74. 7
2. 67. 65. 67. 74.
 78. 61. 60. 72. 62. 74. 66.]
```

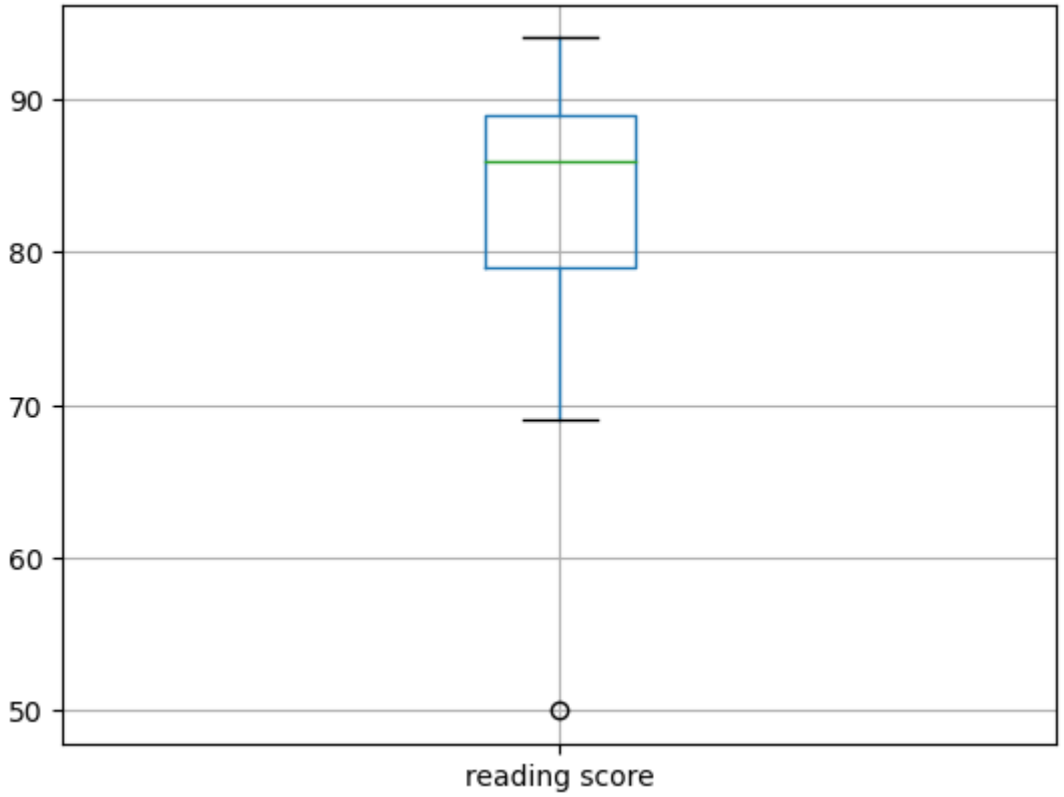
In [46]:
```
df_stud.insert(1,"m score",b,True)
df_stud
```

Out[46]:

	math score	m score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	72.0	86.0	73.0	100.0	2018.0	1.0	Other
1	78.0	78.0	90.0	66.0	79.0	2018.0	2.0	Male
2	77.0	77.0	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.0	73.0	NaN	74.0	88.0	2019.0	3.0	Male
4	74.0	74.0	92.0	90.0	68.0	2018.0	3.0	Other
5	72.0	72.0	94.0	72.0	98.0	2020.0	NaN	Female
7	40.0	40.0	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.0	62.0	78.0	69.0	92.0	2020.0	3.0	Other
9	78.0	78.0	82.0	63.0	86.0	2019.0	3.0	Other
10	77.0	77.0	87.0	66.0	79.0	2019.0	2.0	Male
11	79.0	78.0	NaN	64.0	NaN	2018.0	1.0	Other
12	66.0	66.0	82.0	79.0	93.0	2019.0	3.0	Male
13	74.0	74.0	84.0	69.0	86.0	2020.0	3.0	Male
14	72.0	72.0	89.0	75.0	83.0	2020.0	NaN	Male
17	67.0	67.0	91.0	75.0	87.0	2018.0	3.0	Male
18	65.0	65.0	75.0	NaN	102.0	2019.0	3.0	Other

	math score	m score	reading score	writing score	placement score	club join year	placement offer count	gender
19	67.0	67.0	NaN	64.0	83.0	2021.0	2.0	NaN
21	74.0	74.0	87.0	78.0	NaN	2020.0	1.0	Male
22	89.0	78.0	89.0	66.0	97.0	2019.0	3.0	Female
24	61.0	61.0	89.0	69.0	99.0	2020.0	3.0	Other
25	60.0	60.0	78.0	10.0	82.0	2020.0	2.0	Female
26	72.0	72.0	50.0	80.0	99.0	2021.0	3.0	Female
27	62.0	62.0	76.0	73.0	98.0	2020.0	3.0	Male
28	74.0	74.0	88.0	74.0	94.0	2019.0	3.0	Other

In [50]:
```
col = ['reading score']
df.boxplot(col)
plt.show()
```



In [51]:
```
median=np.median(sorted_rscore)
median
```

Out[51]: 71.5

In [52]:
```
refined_df=df
refined_df['reading score'] = np.where(refined_df['reading score']
```

In [53]:
```
refined_df
```

Out[53]:

	math score	m score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	72.0	71.5	73.0	100.0	2018.0	1.0	Other
1	78.0	78.0	71.5	66.0	79.0	2018.0	2.0	Male

	math score	m score	reading score	writing score	placement score	club join year	placement offer count	gender
2	77.0	77.0	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.0	73.0	NaN	74.0	88.0	2019.0	3.0	Male
4	74.0	74.0	71.5	90.0	68.0	2018.0	3.0	Other
5	72.0	72.0	71.5	72.0	98.0	2020.0	NaN	Female
7	40.0	40.0	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.0	62.0	78.0	69.0	92.0	2020.0	3.0	Other
9	78.0	78.0	82.0	63.0	86.0	2019.0	3.0	Other
10	77.0	77.0	71.5	66.0	79.0	2019.0	2.0	Male
11	79.0	78.0	NaN	64.0	NaN	2018.0	1.0	Other
12	66.0	66.0	82.0	79.0	93.0	2019.0	3.0	Male
13	74.0	74.0	84.0	69.0	86.0	2020.0	3.0	Male
14	72.0	72.0	71.5	75.0	83.0	2020.0	NaN	Male
17	67.0	67.0	71.5	75.0	87.0	2018.0	3.0	Male
18	65.0	65.0	75.0	NaN	102.0	2019.0	3.0	Other
19	67.0	67.0	NaN	64.0	83.0	2021.0	2.0	NaN
21	74.0	74.0	71.5	78.0	NaN	2020.0	1.0	Male
22	89.0	78.0	71.5	66.0	97.0	2019.0	3.0	Female
24	61.0	61.0	71.5	69.0	99.0	2020.0	3.0	Other
25	60.0	60.0	78.0	10.0	82.0	2020.0	2.0	Female
26	72.0	72.0	50.0	80.0	99.0	2021.0	3.0	Female
27	62.0	62.0	76.0	73.0	98.0	2020.0	3.0	Male

In [54]:
```
refined_df['reading score'] = np.where(refined_df['reading score']
```

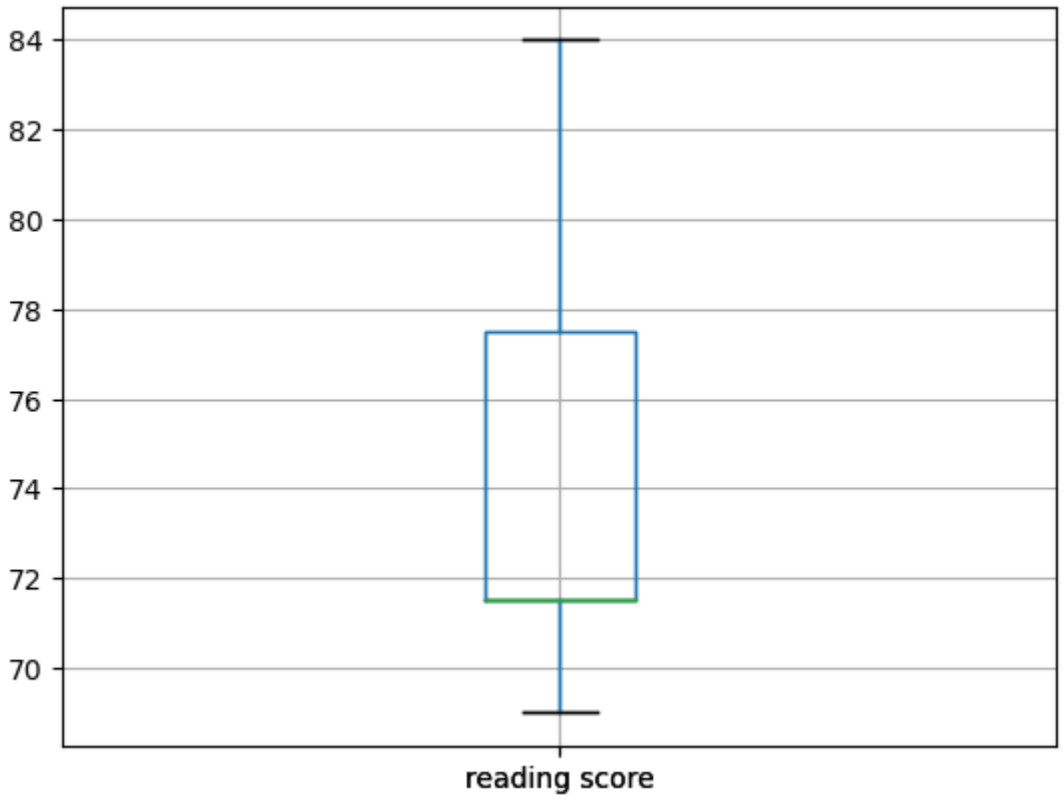
In [56]:
```
refined_df
```

Out[56]:

	math score	m score	reading score	writing score	placement score	club join year	placement offer count	gender
0	72.0	72.0	71.5	73.0	100.0	2018.0	1.0	Other
1	78.0	78.0	71.5	66.0	79.0	2018.0	2.0	Male
2	77.0	77.0	69.0	NaN	NaN	2018.0	1.0	NaN
3	73.0	73.0	NaN	74.0	88.0	2019.0	3.0	Male
4	74.0	74.0	71.5	90.0	68.0	2018.0	3.0	Other
5	72.0	72.0	71.5	72.0	98.0	2020.0	NaN	Female
7	40.0	40.0	82.0	63.0	77.0	2015.0	2.0	NaN
8	62.0	62.0	78.0	69.0	92.0	2020.0	3.0	Other
9	78.0	78.0	82.0	63.0	86.0	2019.0	3.0	Other

	math score	m score	reading score	writing score	placement score	club join year	placement offer count	gender
10	77.0	77.0	71.5	66.0	79.0	2019.0	2.0	Male
11	79.0	78.0	NaN	64.0	NaN	2018.0	1.0	Other
12	66.0	66.0	82.0	79.0	93.0	2019.0	3.0	Male
13	74.0	74.0	84.0	69.0	86.0	2020.0	3.0	Male
14	72.0	72.0	71.5	75.0	83.0	2020.0	NaN	Male
17	67.0	67.0	71.5	75.0	87.0	2018.0	3.0	Male
18	65.0	65.0	75.0	NaN	102.0	2019.0	3.0	Other
19	67.0	67.0	NaN	64.0	83.0	2021.0	2.0	NaN
21	74.0	74.0	71.5	78.0	NaN	2020.0	1.0	Male
22	89.0	78.0	71.5	66.0	97.0	2019.0	3.0	Female
24	61.0	61.0	71.5	69.0	99.0	2020.0	3.0	Other
25	60.0	60.0	78.0	10.0	82.0	2020.0	2.0	Female
26	72.0	72.0	71.5	80.0	99.0	2021.0	3.0	Female
27	62.0	62.0	76.0	73.0	98.0	2020.0	3.0	Male

In [58]:
```
col = ['reading score']
refined_df.boxplot(col)
plt.show()
```

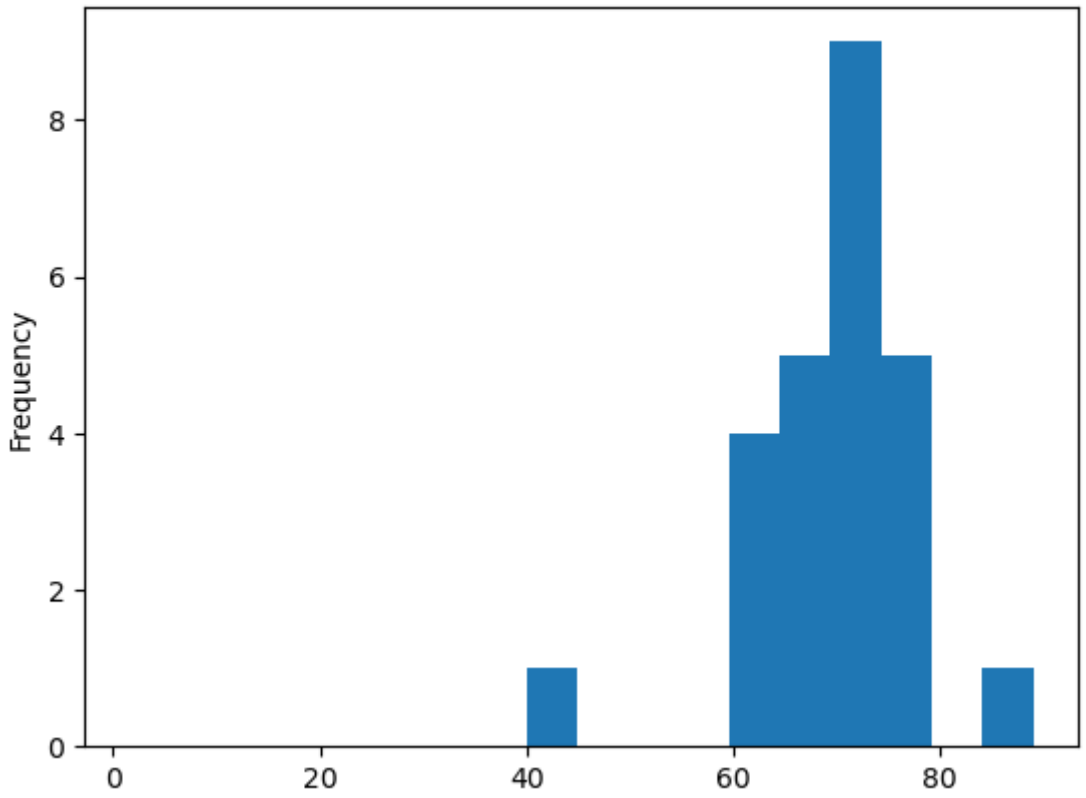


In [59]:
```
new_df['math score'].plot(kind = 'hist')
```

Out[59]: <Axes: ylabel='Frequency'>

In [60]:
```
df['log_math'] = np.log10(df['math score'])
```

In [62]:
```
df['log_math'].plot(kind = 'hist')
plt.show()
```



In [ ]:
```

```